CLIP-Mesh: Generating textured meshes from text using pretrained image-text models

Year: 2022 | Citations: 335 | Authors: N. Khalid, Tianhao Xie, Eugene Belilovsky, T. Popa

Abstract

We present a technique for zero-shot generation of a 3D model using only a target text prompt. Without any 3D supervision our method deforms the control shape of a limit subdivided surface along with its texture map and normal map to obtain a 3D asset that corresponds to the input text prompt and can be easily deployed into games or modeling applications. We rely only on a pre-trained CLIP model that compares the input text prompt with differentiably rendered images of our 3D model. While previous works have focused on stylization or required training of generative models we perform optimization on mesh parameters directly to generate shape, texture or both. To constrain the optimization to produce plausible meshes and textures we introduce a number of techniques using image augmentations and the use of a pretrained prior that generates CLIP image embeddings given a text embedding.